Statistics and Computing (2021) 31:14
https://doi.org/10.1007/s11222-020-09981-3

Bayesian wavelet-packet historical functional linear models

Mark J. Meyer[1] · **Elizabeth J. Malloy**[2] · **Brent A. Coull**[3]

Received: 2 November 2019 / Accepted: 21 October 2020 / Published online: 27 January 2021

Abstract
Historical functional linear models (HFLMs) quantify associations between a functional predictor and functional outcome where the predictor is an exposure variable that occurs before, or at least concurrently with, the outcome. Prior work on the HFLM has largely focused on estimation of a surface that represents a time-varying association between the functional outcome and the functional exposure. This existing work has employed frequentist and spline-based estimation methods, with little attention paid to formal inference or adjustment for multiple testing and no approaches that implement wavelet bases. In this work, we propose a new functional regression model that estimates the time-varying, lagged association between a functional outcome and a functional exposure. Building off of recently developed function-on-function regression methods, the model employs a novel use the wavelet-packet decomposition of the exposure and outcome functions that allows us to strictly enforce the temporal ordering of exposure and outcome, which is not possible with existing wavelet-based functional models. Using a fully Bayesian approach, we conduct formal inference on the time-varying lagged association, while adjusting for multiple testing. We investigate the operating characteristics of our wavelet-packet HFLM and compare them to those of two existing estimation procedures in simulation. We also assess several inference techniques and use the model to analyze data on the impact of lagged exposure to particulate matter finer than 2.5μg, or $PM_{2.5}$, on heart rate variability in a cohort of journeyman boilermakers during the morning of a typical day's shift.

Keywords Functional data analysis · Historical functional models · Bayesian methods and inference · Wavelet-packets · Environmental exposures

1 Introduction

Historical functional linear models (HFLMs) are used to analyze the relationship between a functional "exposure" and a functional "outcome" where only exposures occurring in time before or concurrently with the outcome can affect the outcome. HFLMs are a special case of function-on-function regression (FFR) models, the latter of which fit an unconstrained surface and are therefore inappropriate for modeling functional predictors that are lagged exposures. For example, suppose that for subject i, $x_i(v)$ represents levels of a pollutant sampled on a grid $v \in \mathcal{V}$ and $y_i(t)$ represents a measure of heart health, sampled on a grid $t \in \mathcal{T}$. A general FFR model with no constraints takes the form

$$y_i(t) = \alpha(t) + \int_{v \in \mathcal{V}} x_i(v)\beta(v,t)\mathrm{d}v + E_i(t), \quad (1)$$

where the surface $\beta(v,t)$ is the primary quantity of interest for estimation and $E_i(t)$ is typically assumed to come from a Gaussian process. For example, see Ivanescu et al. (2015), Meyer et al. (2015), Morris (2015), Scheipl et al. (2015), Scheipl and Greven (2016), Kim et al. (2018), and references therein.

This work was supported by Grants from the National Institutes of Health (ES007142, ES000002, ES016454, CA134294). The authors would like to thank Dr. David C. Christiani for use of the journeyman boilermaker data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-020-09981-3.

✉ Mark J. Meyer
mjm556@georgetown.edu

[1] Department of Mathematics and Statistics, Georgetown University, Washington, DC, USA

[2] Department of Mathematics and Statistics, American University, Washington, DC, USA

[3] Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA

However, when applied to data on the association between time-varying personal air pollution exposure and heart rate variability (HRV), model (1) allows HRV measurements at time t to be associated with pollutants occurring both prior to time t as well as after time t, despite the implausibility of such a relationship. The HFLM addresses this issue, constraining $\beta(v,t)$ to prevent such spurious associations by limiting the integration in (1) to the set of coefficients such that $\{v \in \mathcal{V}, t \in \mathcal{T} : v \leq t\}$. The basic HFLM takes the form

$$y_i(t) = \alpha(t) + \int_{\{v \leq t\}} x_i(v)\beta(v,t)\mathrm{d}v + E_i(t). \tag{2}$$

The observed data are usually discrete, so that $\beta(v,t)$ can be expressed as a matrix of coefficients. Thus, the problem reduces to constraining the estimate of $\beta(v,t)$ to be zero for the lower triangle of the matrix.

Several authors explore ways of implementing the constraint in (2). Malfait and Ramsay (2003) propose the use of tent-shaped basis functions with support over a two-dimensional region. They estimate the surface using a multivariate linear model approximation to a finite-dimensional model. After dimension reduction via the basis-space expansion, they use least squares to estimate $\beta(v,t)$. Harezlak et al. (2007) also use basis functions defined over a two-dimensional region, but specify a large number of basis functions and penalize the fit. The authors consider both LASSO and L_2-norm penalties on triangular basis functions, using restricted maximum likelihood (REML) for the latter. Both Malfait and Ramsay (2003) and Harezlak et al. (2007) allow for a pre-defined lag beyond which the effect of exposure is zero, further constraining the surface to a trapezoidal region defined by $\{v, \in \mathcal{V}, t \in \mathcal{T} : t - \Delta \leq v \leq t\}$ for some pre-defined lag Δ. Kim et al. (2011) take the constraint further by proposing a recent history functional linear model where the surface is constrained to a trapezoidal region defined by $\{v, \in \mathcal{V}, t \in \mathcal{T} : t - \Delta_1 \leq v \leq t - \Delta_2\}$ for $0 < \Delta_1 < \Delta_2 < T$. The authors estimate the constrained surface with a varying coefficient model representation using B-spline basis functions, although they suggest Fourier, truncated power, and eigenbasis functions can also be used.

Pomann et al. (2016) and Brockhaus et al. (2017) examine HFLMs with multiple functional predictors and estimation potentially constrained to a fixed window, similar to the work of Harezlak et al. (2007). Pomann et al. (2016) implement two approaches: semi-local smoothing, which performs pointwise estimation, and global smoothing, which smoothes over $\mathcal{T}$ globally. The methods select smoothing parameters via cross-validation and REML, respectively, and use B-spline basis expansions to model the functional form. In each, the covariance of the error term, $E_i(t)$, is assumed independent or to have "working" independence. Brockhaus et al. (2017) use component-wise gradient boosting and B-splines to estimate multiple historical surfaces. To our knowledge, the existing body of work on HFLMs is limited to spline-based methods which can oversmooth signals and peaks in spiky and irregular data. The current literature lacks a Bayesian implementation and also does not consider inferential procedures, focusing instead on estimation and model fit criterion. As such, the performance of the existing methods with respect to uncertainty quantification is not clear.

Wavelet-based functional regression models, such as the work of Morris and Carroll (2006) and Malloy et al. (2010), consider the function-on-scalar and scalar-on-function regression cases, respectively, in the Bayesian context. Meyer et al. (2015) extend the wavelet functional mixed model of Morris and Carroll (2006) to the FFR case using wavelets for the basis function of the outcome and wavelet principal components (WPCs) for reducing the dimension of the predictor. One advantage of the wavelet-based framework is that it does not require the assumption of independence in the data space; that is, the wavelet-based approach can accommodate a wide range of structures for the covariance of $E_i(t)$ (Morris and Carroll 2006). However, the wavelet-based FFR of Meyer et al. (2015) fits an unconstrained surface, as in (1), and cannot be used to estimate historical effects, as in (2). To implement an HFLM using wavelets, we need a wavelet transformation that maintains the temporal ordering of the untransformed outcome and exposure data. The discrete wavelet transformation (DWT) and WPC, which Meyer et al. (2015) employ, distort the original time domains and thus do not maintain the exposure–response ordering. However, the discrete wavelet-packet transformation (DWPT), which has not previously been used for functional regression, is an alternative algorithm for determining wavelet-space coefficients. The indexing of the resulting coefficients has a convenient relationship to the time domain that we can exploit, allowing us to sample coefficients in such a way that maintains the historical constraint in the wavelet-space.

In this work, we propose a novel use of the DWPT to, in conjunction with Bayesian regularization, enforce the constraint in model (2). The resulting coefficients from a DWPT maintain the ordering of the original data. Thus, after performing the DWPT on both the y and x, the coefficients are similarly ordered. While the DWPT alone does not constrain the surface, its preservation of the temporal ordering of the original signal in the wavelet coefficients allows us to use a prior in the wavelet-space to constrain the surface and thus build a Bayesian wavelet-space HFLM. We formulate our model within the framework of Morris and Carroll (2006) and Meyer et al. (2015). Thus, our method does not rely on the assumption of independence in the data space and can accommodate a wide range of underlying within-curve covariance structures. A benefit of the Bayesian approach is that we can implement joint credible intervals to quantify uncertainty and Simultaneous Band Scores or SimBaS

(Meyer et al. 2015) to identify critical regions of time when the exposure is associated with the outcome—both of which adjust for multiplicity. We assess the operating characteristics of the methodology in simulation and show that, in comparison with two other existing approaches for HFLMs, our method performs well. We also demonstrate the advantages of using joint intervals and SimBaS over other Bayesian methods for inference including point-wise credible intervals and the Bayesian false discovery rate. Finally, we present an application to data from a study of journeyman boilermakers exposed to particulate matter during the course of a workday.

The motivating data consist of five-minute assessments of HRV, as defined by the standard deviation of the normal-to-normal intervals (SDNN) at each time point, and particulate matter finer than 2.5 μm ($PM_{2.5}$) which results from exposure to residual oil fly ash and cigarette smoke (Magari et al. 2001; Cavallari et al. 2008). Harezlak et al. (2007) present an analysis of part of this data and found both negative and positive time-specific associations in the morning that corresponded to the workers' break times. Our analysis focuses on the morning hours of the workday where Harezlak et al. (2007) demonstrate the largest effects. We make available MATLAB code for the implementation of our method at https://github.com/markjmeyer/WPHFLM.

The remainder of the paper is organized as follows. Section 2 provides a brief introduction to the discrete wavelet-packet transformation. Section 3 details the formulation of the Bayesian wavelet-packet HFLM along with a discussion of inferential procedures. Sections 4 and 5 present the results of our simulation study and the application of our model to the journeyman data, respectively. Finally, in Sect. 6, we provide a discussion of the methodology.

2 Discrete wavelet-packet transformation

Consider a one-dimensional function, $x(v)$, which we discretely observe as $\mathbf{x} = [x_1 \cdots x_V]'$, where $V = 2^k$ for some positive integer value k (wavelet transformations require the signal to be dyadic, see Percival and Walden (2000) and Nason (2008) for more details). For a given mother wavelet, the DWPT begins by first passing the corresponding low-, $\mathcal{G}$, and then high-pass, $\mathcal{H}$, filters over the original signal, resulting in a set of approximation coefficients, $\mathbf{a}$, and detail coefficients, $\mathbf{d}$. After the first level, the DWPT applies $\mathcal{G}$ and $\mathcal{H}$ to all coefficients within the level, reversing the order when applied to the detail coefficients. Figure 1 illustrates the partial DWPT decomposition for $J = 3$ levels. The resulting representation of $\mathbf{x}$ is the set of coefficients from the last level, $\mathbf{w}_P = [\mathbf{a}_{3,0}\ \mathbf{d}_{3,1}\ \mathbf{a}_{3,2}\ \mathbf{d}_{3,3}\ \ \mathbf{a}_{3,4}\ \mathbf{d}_{3,5}\ \mathbf{a}_{3,6}\ \mathbf{d}_{3,7}]'$. The DWPT can be expressed as a post-multiplication of the original exposure vector by an orthogonal matrix, W_P. Thus, $\mathbf{w}_P$ can be shown to be $\mathbf{w}_P = \mathbf{x}W_P$.

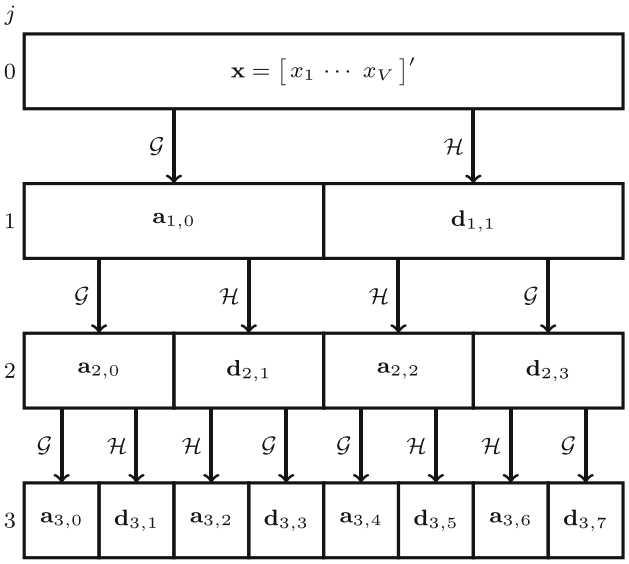

Fig. 1 Algorithm for partial, $j = 3$, DWPT (Percival and Walden 2000)

Similar to the DWT, the wavelet-packet coefficients in $\mathbf{w}_P$ are indexed by a scale and location. The scale indexes the bin the coefficient is in at the final level of the decomposition and thus corresponds to the second subscript from the $j = 3$ level in Fig. 1, while the location denotes the position of the coefficient within the set of coefficients at a given scale. Thus, the time ordering of the observed signal is preserved, within a scale, in the ordering of the coefficients. Suppose we have a second function, $y(t)$, that we discretely observe on a grid such that $\mathbf{y} = [y_1 \cdots y_T]'$. Performing the DWPT on y, we obtain a set of wavelet-packet coefficients that preserve the time ordering of the original signal within each scale. Provided the elements of $\mathbf{x}$ are sampled in time concurrently or before the elements of $\mathbf{y}$, we can use the location index from their respective DWPT decompositions to estimate only the desired region of the surface within each scale in the wavelet-packet space. It is important to note that the DWPT does not, on its own, constrain the surface. It only provides a convenient representation of the signal in the wavelet-packet space that allows us to sample only certain coefficients. When we apply the inverse DWPT (IDWPT), the constraint is maintained in the data space. For more details on DWPTs, see Percival and Walden (2000, chap. 6), Misiti et al. (2007), and Nason (2008, chap. 2).

3 Bayesian historical functional linear model

We begin with the model in (2) that constrains the estimation to the region defined by $\{v \in \mathcal{V}, t \in \mathcal{T} : v \leq t\}$. We assume the within-function errors come from a Gaussian process, $E_i(t) \sim \mathcal{GP}(0, \Sigma_E)$, where Σ_E is an unstructured covariance matrix. Because the data, $y_i(t)$ and $x_i(v)$, arrive sampled

on a grid of equally spaced time points $t = [t_1, \ldots, t_T]'$ and $v = [v_1, \ldots, v_V]'$, we use the vector version of the model: $\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{e}_i$ for the vectors $\mathbf{y}_i$, $\mathbf{x}_i$, and $\mathbf{e}_i$ and matrix of coefficients $\boldsymbol{\beta}$. We recommend centering and scaling both the outcome and predictor functions first. Thus, without loss of generality, we drop the intercept function from the model formulation—although our method does estimate the intercept and can accommodate scalar covariates. Stacking the response vectors and predictor vectors into matrices gives

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \tag{3}$$

where for N total curves, $\mathbf{E}$ and $\mathbf{Y}$ are $N \times T$, while $\mathbf{X}$ is $N \times V$. The constrained region of integration in Model (2) restricts the form of the functional regression coefficients so that $\beta(v_k, t_{k'}) = 0$ if $v_k > t_{k'}$. If $T = V$ and $t_1 = v_1, t_2 = v_2, \ldots, t_T = v_V$, then the discrete version of $\boldsymbol{\beta}$ is an upper triangular matrix of the form

$$\boldsymbol{\beta} = \begin{pmatrix} \beta(v_1,t_1) & \beta(v_1,t_2) & \cdots & \beta(v_1,t_T) \\ 0 & \beta(v_2,t_2) & \cdots & \beta(v_2,t_T) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta(v_V,t_T) \end{pmatrix}, \tag{4}$$

with zeros below the main diagonal.

3.1 Historical constraint via wavelet-packets

Working with (3), we apply the DWPT separately to each row of $\mathbf{Y}$ and to each row of $\mathbf{X}$. Performing this transformation is equivalent to the post-multiplication of the approximately orthonormal projection matrices resulting from the DWPT (Percival and Walden 2000). The decompositions have the form $\mathbf{Y} = \mathbf{Y}^{W_P}\mathbf{W}_{P,Y}$ and $\mathbf{X} = \mathbf{X}^{W_P}\mathbf{W}_{P,X}$ where $\mathbf{W}_{P,Y}$ and $\mathbf{W}_{P,X}$ are orthogonal matrices containing the wavelet-packet basis functions. Then, for the two-dimensional decomposition on $\boldsymbol{\beta} = \mathbf{W}'_{P,X}\boldsymbol{\beta}^{W_P}\mathbf{W}_{P,Y}$, Model (3) in the wavelet-packet space is $\mathbf{Y}^{W_P}\mathbf{W}_{P,Y} = \mathbf{X}^{W_P}\mathbf{W}_{P,X}\mathbf{W}'_{P,X}\boldsymbol{\beta}^{W_P}\mathbf{W}_{P,Y} + \mathbf{E}^{W_P}\mathbf{W}_{P,Y}$ for $\mathbf{E} = \mathbf{E}^{W_P}\mathbf{W}_{P,Y}$. Post-multiplying by $\mathbf{W}_{P,Y}$ and recognizing the orthogonality of the wavelet-packet basis matrices, this model reduces to $\mathbf{Y}^{W_P} = \mathbf{X}^{W_P}\boldsymbol{\beta}^{W_P} + \mathbf{E}^{W_P}$ with subject-specific model $\mathbf{y}_i^{W_P} = \mathbf{x}_i^{W_P}\boldsymbol{\beta}^{W_P} + \mathbf{e}_i^{W_P}$.

We enforce the constraint in the wavelet-packet space via our prior specification on the elements of $\boldsymbol{\beta}$. Let the DWP transformations be indexed by scales $j = 1, \ldots, J^y$ and $s = 1, \ldots, S^x$ and locations $k = 1, \ldots, K^y$ and $\ell = 1, \ldots, L^x$ in the $\mathbf{Y}$ and $\mathbf{X}$ wavelet-packet spaces, respectively. Consistent with previous work on wavelet-based models in function regression, we place spike-and-slab priors on model coefficients. To restrict the surface in the wavelet-packet space, our prior on the elements of $\boldsymbol{\beta}^{W_P} = \left[\beta_{s\ell,jk}^{W_P}\right]$ is $\beta_{s\ell,jk}^{W_P} \sim 1(\ell \le k)\gamma_{s\ell,jk}N(0, \tau_{s\ell,j\cdot}) + [1 - \gamma_{s\ell,jk}]d_0$, where $\gamma_{s\ell,jk} \sim Bern(\pi_{s\ell,j\cdot})$ and d_0 is a point mass distribution at zero. The regularization parameters, $\tau_{s\ell,j\cdot}$ and $\pi_{s\ell,j\cdot}$, smooth over locations k which we denote using the "dot" notation in the subscript. We assume inverse gamma and beta hyper-priors, respectively, for the regularization parameters with hyper-parameters based on empirical Bayes estimates.

Morris and Carroll (2006) show that after a wavelet transformation, assuming independence in the wavelet-space does not imply independence in the data space. Therefore, wavelets accommodate a wide range of covariances in the data space. As wavelet-packets share the same whitening properties of wavelets, we assume independence in the wavelet-packet space (Percival and Walden 2000). Thus, we let $\mathbf{e}_i^{W_P} \sim N(0, \boldsymbol{\Sigma}^{W_P})$ where $\boldsymbol{\Sigma}^{W_P} = \text{diag}\left\{\sigma_{jk}^2\right\}$, which varies by the scale and location of the $\mathbf{Y}$ wavelet-packet coefficients. We place an inverse gamma prior on σ_{jk}^2. The independence assumption allows us to sample the coefficients separately, corresponding to different combinations of j and k .

For the jkth wavelet-packet space coefficient from $\mathbf{Y}^{W_P}$ and the $s\ell$th column of $\mathbf{X}^{W_P}$, the conditional posterior distribution is a mixture of a point mass at zero and a normal distribution of the form

$$\begin{aligned}\beta_{s\ell,jk}^{W_P}|\text{rest} \sim & 1(\ell \le k)\gamma_{s\ell,jk}N(\mu_{s\ell,jk}, \epsilon_{s\ell,jk}) \\ & + (1 - \gamma_{s\ell,jk})d_0,\end{aligned} \tag{5}$$

where $\mu_{s\ell,jk} = \hat{\beta}_{s\ell,jk}^{W_P}(1 + \Lambda_{s\ell,jk}/\tau_{s\ell,j})^{-1}$ and $\epsilon_{s\ell,jk} = \Lambda_{s\ell,jk}(1 + \Lambda_{s\ell,jk}/\tau_{s\ell,j})^{-1}$ for the OLS and variance estimates $\hat{\beta}_{s\ell,jk}^{W_P}$ and $\Lambda_{s\ell,jk}$ at the current step. The conditional for $\gamma_{s\ell,jk}$ is

$$\gamma_{s\ell,jk}|\text{rest} \sim Bern(\alpha_{s\ell,jk}), \tag{6}$$

where $\alpha_{s\ell,jk} = O_{s\ell,jk}/\left(O_{s\ell,jk} + 1\right)$ for

$$O_{s\ell,jk} = \pi_{s\ell,j}/(1 - \pi_{s\ell,j})\text{BF}_{s\ell,jk},$$

$$\text{BF}_{s\ell,jk} = e^{\frac{1}{2}\zeta_{s\ell,jk}^2\left(1+\frac{\Lambda_{s\ell,jk}}{\tau_{s\ell,j}}\right)} \Big/ \sqrt{1 + \frac{\tau_{s\ell,j}}{\Lambda_{s\ell,jk}}},$$

and $\zeta_{s\ell,jk}$ equal to the ratio of the current values of $\beta_{s\ell,jk}^{W_P}$ to the current estimate of the standard deviation of $\beta_{s\ell,jk}^{W_P}$. Via the indicator function in the conditional in (5), $1(\ell \le k)$, we enforce the historical constraint by forcing coefficients for which $\ell > k$ to come from the point mass density d_0 and thus be constrained to zero. Because of wavelet-packet space independence, these coefficients do not inform the estimation of the coefficients in the region of interest and are excluded from informing the variance components.

The conditionals for the diagonal elements of the wavelet-packet space variance components, σ^2_{jk}, are

$$P\left(\sigma^2_{jk}|\text{rest}\right) \propto \pi\left(\sigma^2_{jk}\right)\left(\sigma^2_{jk}\right)^{-n/2} \times \exp\left[-\frac{1}{2\sigma^2_{jk}}\left(\mathbf{y}^{W_P}_{jk} - X\beta^{W_P}_{\cdot\cdot,jk}\right)'\left(\mathbf{y}^{W_P}_{jk} - X\beta^{W_P}_{\cdot\cdot,jk}\right)\right], \quad (7)$$

where $\pi\left(\sigma^2_{jk}\right)$ is the prior density on σ^2_{jk}, which we take to be inverse gamma with parameters a_{σ^2} and b_{σ^2} both set to their empirical Bayes estimates. Similar to prior work on wavelet-based functional regression models, we employ a Metropolis–Hastings step to sample these variance components (Morris and Carroll 2006; Morris et al. 2008; Malloy et al. 2010; Zhu et al. 2011; Meyer et al. 2015). The proposal densities are independent Gaussians, truncated at zero and centered at the previous value in each chain. The full conditionals for the regulationization parameters are

$$\tau_{s\ell,j\cdot}|\text{rest} \sim IG\left[a_\tau + \frac{1}{2}\gamma_{s\ell,jk},\ b_\tau + \frac{1}{2}\gamma_{s\ell,jk}\left(\beta^{W_P}_{s\ell,jk}\right)^2\right] \text{ and} \quad (8)$$

$$\pi_{s\ell,j\cdot}|\text{rest} \sim \text{Beta}\left(a_\pi + \gamma_{s\ell,jk}, b_\pi + \gamma_{s\ell,jk}\right), \quad (9)$$

with a_τ, b_τ, a_π, and b_π set to the empirical Bayes estimates. These prior specifications are in accordance with the previous work on wavelet-based functional models (Morris and Carroll 2006; Malloy et al. 2010; Meyer et al. 2015). For more details on the empirical Bayes estimates, see Morris and Carroll (2006). Our sampler then iterates between draws from (5) to (9) until convergence. Upon completion of the algorithm, we apply the inverse DWPT to the posterior samples of $\boldsymbol{\beta}^{W_P}$ to obtain estimates in the data space of $\boldsymbol{\beta}$, the upper triangular matrix of historically constrained coefficients.

3.2 Thresholding and wavelet details

There are two potential sources of computational burden in the estimation procedure. The first source is in the post-processing of the samples when projecting them back into the data space which requires pre- and post-multiplication of each sampled matrix. For several of the inference procedures we consider, the full posterior sample must be obtained and this inefficiency is unavoidable. However, if the estimate and point-wise credible intervals are all that are of interest, the wavelet-space samples can be summarized accordingly and then projected into the data space. This considerably reduces computation time. The second source is the sampler itself which becomes computationally intensive as V increases. In previous work on wavelet-based models, Meyer et al. (2015) address this by reducing the dimension of the wavelet transformed $\mathbf{X}$ using wavelet principal components (WPC) and retaining columns containing a large amount of the variability in $\mathbf{X}$. The WPC decomposition involves first performing a DWT on $\mathbf{X}$ and then performing a singular value decomposition (SVD). While this approach reduces computation time and achieves additional denoising akin to thresholding coefficients to zero, it does not work in the historical framework because performing an SVD on $\mathbf{X}^{W_P}$ disrupts the temporal ordering.

To address the computational intensity of the sampler and simultaneously achieve additional denoising as is commonly done with wavelets via hard or soft thresholding, we propose a simplified thresholding procedure: All coefficients in the larger scales of $\mathbf{X}^{W_P}$ are set to zero. We consider retaining the scales that comprise 25 and 50% of the wavelet-packet coefficients in $\mathbf{X}^{W_P}$. For example, using the $J = 3$ level partial DWPT, the first approach would result in retaining the vectors of wavelet-packet coefficients $\mathbf{a}_{3,0}$ and $\mathbf{d}_{3,1}$ from Fig. 1. The second approach, retaining 50%, would correspond to retaining the vectors of wavelet-packet coefficients $\mathbf{a}_{3,0}$, $\mathbf{d}_{3,1}$, $\mathbf{a}_{3,2}$, and $\mathbf{d}_{3,3}$, also from Fig. 1. Our simulation compares the performance of both approaches in terms of computational efficiency and estimation. It is important to note that we do not threshold the outcome and retain all coefficients of $\mathbf{Y}^{W_P}$.

For the choice of mother wavelets, we use Daubechies wavelets with three vanishing moments for the separate DWPTs on both $\mathbf{Y}$ and $\mathbf{X}$. Wavelet-packets also require a choice of boundary padding for most types of mother wavelets and when the signal is not dyadic. We select zero padding, which pads each boundary of the vector of coefficients with zeros (Percival and Walden 2000). To ensure the padding is consistent within each scale between $\mathbf{Y}$ and $\mathbf{X}$, we wrote our own DWPT algorithm that applies the padding uniformly within each scale. Finally, we must choose the number of levels for the partial DWPT. Because we seek to maintain time ordering within scales, we must use the same number of levels of decomposition, $J = 3$, for both $\mathbf{Y}$ and $\mathbf{X}$. We consider the impact of using $J = 4$ levels on the journeyman data in the application.

3.3 Posterior inference and model evaluation

Previous implementations of the HFLM have focused on model fit, and not on statistical inference. In the wavelet-based functional literature, Meyer et al. (2015) use a Bayesian false discovery rate (BFDR) procedure that Morris et al. (2008) and Malloy et al. (2010) also consider. Meyer et al. (2015) also use joint credible bands (Ruppert et al. 2003) and propose Simultaneous Band Scores or SimBaS. We consider the use of each method in the context of Bayesian wavelet-packet HFLMs.

The BFDR utilizes the MCMC samples to estimate the posterior probability of a coefficient being greater than a meaningful effect size, δ. These values are ranked and a cut-off selected to control the overall FDR at a pre-specified global α-bound. Suppose we have M MCMC samples and $\beta^{(m)}(v,t)$ is mth draw from the posterior estimated surface. For $\{v \in \mathcal{V}$ and $t \in \mathcal{T}$ s.t. $v \leq t\}$, we find $P_B(v,t) = Pr\{|\beta(v,t)| > \delta | y\} \approx \frac{1}{M}\sum_{m=1}^{M} 1\left\{\left|\beta^{(m)}(v,t)\right| > \delta\right\}$. We then flag the set of coefficients on the historical surface that satisfy $\psi = \{(v,t) : v \leq t$ and $P_B(v,t) \geq \phi_\alpha\}$. Here, ϕ_α is determined by ranking the values of P_B in descending order to obtain the set $\left\{P_{(r)}, r = 1, \ldots, R\right\}$, where R is the number of coefficients satisfying the historical constraint. Then, define the cutoff value as λ, $\lambda = \max\left[r^* : \frac{1}{r^*}\sum_{r=1}^{r^*}\left\{1 - P_{(r)}\right\} \leq \alpha\right]$. We select coefficients with P_B greater than or equal to $\phi_\alpha = P_{(\lambda)}$ as significant.

For interval estimation, we consider both point-wise credible intervals (PWCIs) as well as joint credible intervals. PWCIs are constructed by finding the $\alpha/2$ and $1-\alpha/2$ quantiles of the posterior samples taken at each coefficient for some choice of α. Similar to Meyer et al. (2015), we construct joint credible intervals using $I_\alpha(v,t) = \hat{\beta}(v,t) \pm q_{(1-\alpha)}\left[\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}\right]$, where $\hat{\beta}(v,t)$ and $\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}$ are the mean and standard deviation of the posterior samples, respectively, and $q_{(1-\alpha)}$ is the $(1-\alpha)$ quantile taken over all posterior samples of the quantity

$$q^{(m)} = \max_{(v,t)}\left|\frac{\beta^{(m)}(v,t) - \hat{\beta}(v,t)}{\widehat{\text{St.Dev}}\left\{\hat{\beta}(v,t)\right\}}\right|, \text{ s.t. } v \leq t.$$

Such an interval satisfies $Pr\{L(v,t) \leq \beta(v,t) \leq U(v,t)$ $\forall\, v \in \mathcal{V}, t \in \mathcal{T}$ s.t. $v \leq t\} \geq 1-\alpha$, where $L(v,t)$ and $U(v,t)$ are the corresponding upper and lower interval bounds. This procedure yields a joint $100(1-\alpha)$ interval for the historical association surface. SimBa scores are then defined as the smallest value of α at which the interval first excludes zero for each (v,t).

For model selection, Harezlak et al. (2007) and Meyer et al. (2015) both propose the use of a functional R^2 metric of the form

$$R^2_{\text{ave}} = \frac{1}{T}\int_0^T\left[1 - \frac{\sum_{i=1}^{n}\{y_i(t) - \hat{y}_i(t)\}^2}{\sum_{i=1}^{n}\{y_i(t)\}^2}\right]\text{d}t.$$

In the application, we use this metric for selecting between retaining 25% and 50% of the wavelet-packet coefficients in $\mathbf{X}^{W_P}$. Thus, we apply R^2_{ave} to the wavelet-space model:

$$R^2_{W,ave} \approx \frac{1}{J^y K^y}\sum_j\sum_k\left[1 - \frac{\sum_{i=1}^{n}\{\mathbf{y}^{W_P}_{i,jk} - \hat{\mathbf{y}}^{W_P}_{i,jk}\}^2}{\sum_{i=1}^{n}\{\mathbf{y}^{W_P}_{i,jk}\}^2}\right].$$

4 Simulation study

We first consider a setting designed to mimic our application with $N = 20$ and $T = V = 32$ which we refer to as the peak effect. The top left panel of Fig. 2 displays the peak surface. We then examine four other historical surfaces representing more general relationships between $x(v)$ and $y(t)$: a lagged effect of $x(v)$ on $y(t)$, a cumulative effect, a time-specific effect, and a delayed time-specific effect. Graphs of these four more general surfaces along with the corresponding mathematical expressions for all surfaces are in Supplementary Material. For each of the more general relationships, we vary the sampling rate such that $T = V = 64$ and $T = V = 128$. We also vary the sample size, considering $N = 50$ and $N = 200$. For the peak and the sparser, $T = 64$ settings, we vary the percent of the columns of $\mathbf{X}^{W_P}$ we retain: first retaining 25% of coefficients and then retaining 50%. In the denser, $T = 128$ settings, we only retain 25% of coefficients since retaining more considerably increases the computational burden. For comparison, we also implement the FDBoost approach proposed by Brockhaus et al. (2017) using 1500 boosting iterations and the finite element basis (FEB) approach used by Malfait and Ramsay (2003) using 13 basis functions, per the authors' recommendation.

For each "true" historical surface $\boldsymbol{\beta}$, we generate N $\mathbf{x}_i$ curves from a mean zero Gaussian Process with a first-order auto-regressive (AR1) covariance structure. We base the variance and correlation parameters of the AR1 covariance off of the $PM_{2.5}$ data from the journeyman data, letting $\sigma^2_{AR,X} = 3.5$ and $\rho_X = 0.75$. Next we generate within-subject error functions, $\mathbf{e}_i$, from a separate mean zero Gaussian Process with an AR1 covariance structure. Once again we base the parameters of the covariance matrix off of the journeyman data, setting them to $\sigma^2_{AR,E} = 0.1$ and $\rho_E = 0.5$. We simulate the outcome functions using $\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{e}_i$. We repeat this data generation process 200 times for all settings, each time obtaining 2000 posterior samples, discarding the first 1000. For every simulated dataset, we perform a $J = 3$ level DWPT on both the outcome and predictor using Daubechies wavelets with 3 vanishing moments. All computation is done using MATLAB version R2017a on a desktop with a 3.2 GHz Intel Core i5 processor and 16 GB of memory.

To evaluate each method's performance in estimation, we find the root mean integrated squared error (RMISE). Table 1 presents the RMISE averaged across all 200 simulated datasets for the peak scenario, while Fig. 2 shows a single estimated surface with near-average RMISE from the wavelet-packet model retaining 25% of coefficients in the top right panel. Table 2 presents RMISEs for the more general relationships, while corresponding figures similar to those in Fig. 2 are in Supplementary Material.

From Table 1, the wavelet-packet model retaining 25% of coefficients has the smallest average RMISE, while FDBoost

Table 1 RMISE for the peak setting with $N = 20$ and $T = 32$ for the wavelet-packet model retaining 25% (WP 25%) and 50% of coefficients (WP 50%) alongside the FDBoost and finite element basis (FEB) approaches

	RMISE	Coverage	
		Point-wise	Joint
WP 25%	0.044	0.780	0.943
WP 50%	0.073	0.827	0.973
FEB	0.106	–	–
FDBoost	0.062	–	–

Coverage for point-wise and joint credible intervals from the wavelet-packet models is in the last two columns

is the next smallest. The wavelet-packet model retaining 50% of coefficients is slightly larger than FDboost while the FEB's RMISE is an order of magnitude larger. In Table 2, the wavelet-packet models outperform both FDBoost and FEB across the board. Increasing the sample size tends to decrease RMISE for the wavelet-packet models as does increasing T and V. The more complicated relationships, $\boldsymbol{\beta}_T$ and $\boldsymbol{\beta}_D$, have lower RMISEs when retaining more coefficients, although all settings produce similar results. Regardless of sample size or sampling density, retaining 25% of the X-space wavelet-packet coefficients preserves roughly 75% of the energy, while retaining 50% of coefficients preserves 88% of the energy in $X(v)$. To calculate average preserved energy, we use $\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i^{Wp\prime}\mathbf{x}_i^{Wp}$ from the thresholded wavelet-packet coefficients. Figure 2 suggests that a single simulated dataset with near-average RMISE accurately estimates the peak effect, even when only retaining an average of 75% of the energy in the predictor function. Results under the more general relationships are similar; see Supplementary Material.

Computation time comes from two sources: the sampler itself and the post-processing of the samples. For the peak

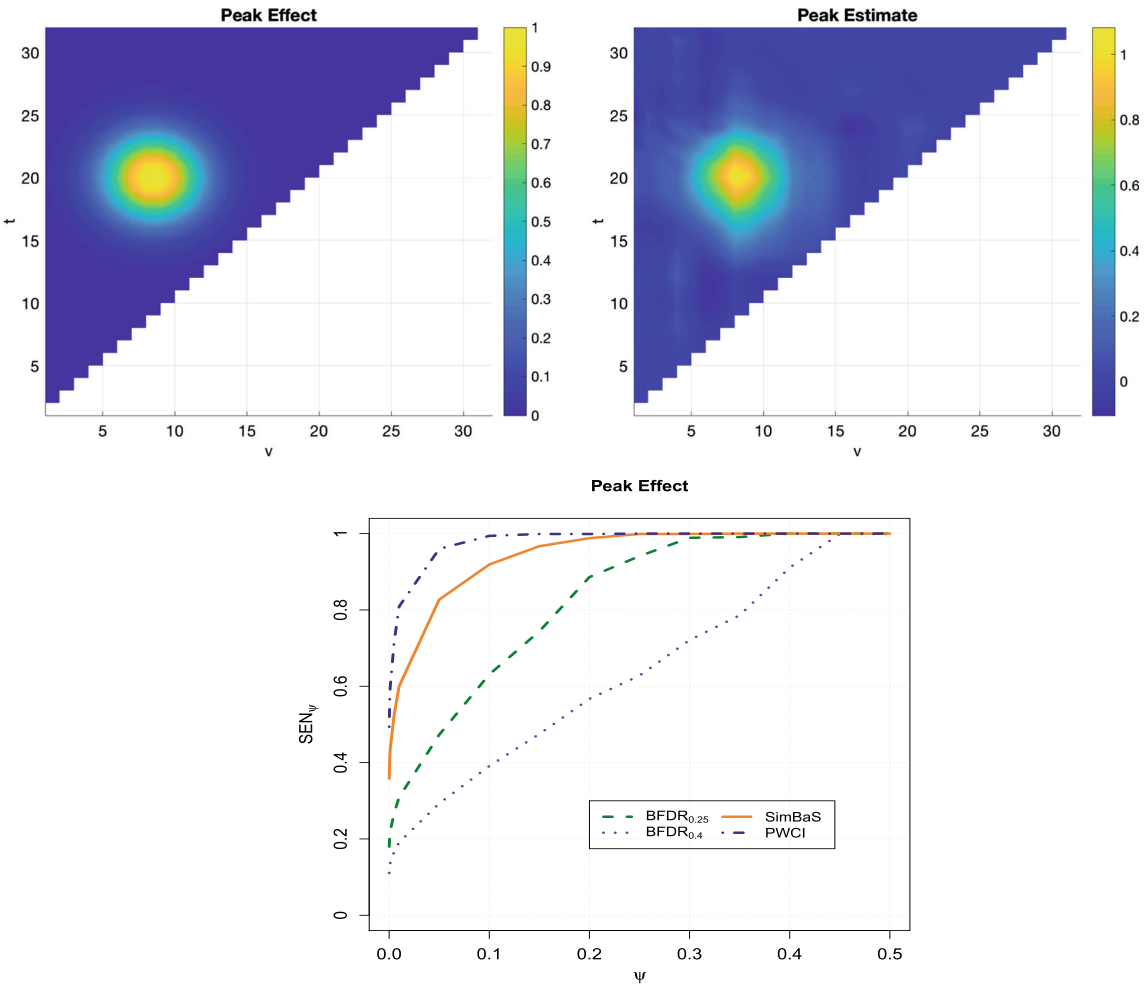

Fig. 2 True peak effect (upper left panel) and a single estimated surface with near-average RMISE (upper right panel). The sensitivity, SEN$_\psi$, for the SimBaS, point-wise credible intervals (PWCIs), and BFDRs at $\delta = 0.25$ (BFDR$_{0.25}$) and 0.4 (BFDR$_{0.4}$). This figure appears in color in the electronic version of this article

Table 2 RMISEs for the general relationships lagged (L), cumulative (C), time-specific (T), and delayed time-specific (D)

N	T	Model	RMISE			
			β_L	β_C	β_T	β_D
50	64	WP 25%	0.015	0.012	0.095	0.045
		WP 50%	0.018	0.017	0.056	0.017
		FEB	0.097	0.072	0.259	0.078
		FDBoost	0.064	0.040	0.296	0.048
	128	WP 25%	0.010	0.008	0.040	0.008
		FEB	0.099	0.079	0.267	0.085
		FDBoost	0.055	0.037	0.284	0.022
200	64	WP 25%	0.009	0.007	0.088	0.009
		WP 50%	0.010	0.008	0.045	0.008
		FEB	0.097	0.071	0.282	0.077
		FDBoost	0.055	0.034	0.296	0.028
	128	WP 25%	0.004	0.004	0.036	0.003
		FEB	0.100	0.079	0.278	0.085
		FDBoost	0.049	0.034	0.241	0.023

Table values represent averages taken over 200 simulated datasets. RC denotes retained coefficients. WP 25% and WP 50% denote the wavelet-packet models retaining 25% and 50% of coefficients, respectively. FEB denotes the finite element basis approach

setting, the sampler takes an average of 12.2 s when retaining 25% of the coefficients and 17.5 s when retaining 50%. Regardless of retained coefficients, post-processing takes an average of 262.6 s for a total computation time between 274.8 and 280.1 s. As V increases, we see greater gains in computational efficiency resulting from the reduction. When $V = 64$, models retaining 25% of the coefficients take an average of 59.7 s for the sampler to finish. The sampler for models retaining 50% of the coefficients finishes, on average, in 158.2 s. The post-processing is similar with both models taking an average of 455.3 s making the total computation time between 515.9 and 612.5 s, depending on the number of retained coefficients. For $T = V = 128$, the models take longer with the sampler completing in 195.4 s and post-processing taking 838.1 s for a total of 1033.5 s, on average.

We assess inference using coverage along with metrics defined in Meyer et al. (2015) for assessing sensitivity and false discovery rate in functional regression. The first metric, denoted SEN_Ψ, determines the proportion of coefficients selected as significant out of the number of coefficients from the true surface that are greater than Ψ. The second metric, FDR_ϵ, is the number of coefficients flagged as significant whose true value is $\leq \epsilon$ divided by the number of flagged coefficients. Coverage for both point-wise and joint credible intervals under the peak setting is given in Table 1, while Table 3 contains coverage probabilities for the general relationships. Coverage is calculated by first averaging over the historical surface and then over all 200 simulated datasets.

In general, we see that the joint credible intervals provide higher coverage. In the peak setting, the joint interval from the model retaining 25% of coefficients attains average coverage that is closest to nominal. Coverage also tends to increase, regardless of interval type, as the percent of retained coefficients increase, though in some settings the differences are negligible. Coverage of the joint intervals can also be affected by sampling density and sample size. However, for the joint intervals, coverage is above the nominal level for most settings. This is not the case for the PWCI, where coverage sits below the nominal level—in some instances considerably below it. All intervals are at the 95% level.

Given the similarity in the RMISEs, we assess sensitivity and FDR of the inference procedures for only the wavelet-packet model that retains 25% of the coefficients. For this model applied to the peak setting, the bottom panel of Fig. 2 presents the sensitivity as a function of Ψ, while Table 4 displays FDR_ϵ values. Sensitivity for the general relationships under a sample size of $N = 50$ and sampling density of $T = 64$ are in Fig. 3. Similar figures for the remaining combinations along with tables of FDR_ϵ for all settings not displayed here are in Supplementary Material. For the BFDR, we must select a δ value which represents a meaningful change in the effect. Since the outcome function is scaled prior to analysis, we investigate δ's of 0.25 and 0.4 which represent one-quarter and two-fifths of a standard deviation change in $y_i(t)$, respectively, per one standard deviation change in $x_i(t)$.

Table 3 Point-wise and joint credible interval coverage probabilities for the simulation settings lagged (L), cumulative (C), time-specific (T), and delayed time-specific (D)

N	T	RC (%)	Point-wise interval				Joint interval			
			β_L	β_C	β_T	β_D	β_L	β_C	β_T	β_D
50	64	25	0.811	0.870	0.524	0.565	0.993	0.997	0.770	0.778
		50	0.923	0.942	0.826	0.934	0.999	0.999	0.984	0.999
	128	25	0.855	0.885	0.767	0.940	1.000	1.000	0.978	1.000
200	64	25	0.724	0.784	0.452	0.735	0.957	0.986	0.673	0.957
		50	0.865	0.890	0.718	0.930	0.999	0.999	0.933	1.000
	128	25	0.884	0.881	0.590	0.897	0.999	0.999	0.906	1.000

All intervals are at the 95% level. Table values represent averages taken over 200 simulated datasets. RC denotes retained coefficients

Table 4 False discovery rates, FDR_ϵ, for the peak setting with $N = 20$ and $T = 32$ for SimBaS, point-wise credible intervals (PWCIs), and BFDRs at $\delta = 0.25$ ($BFDR_{0.25}$) and 0.4 ($BFDR_{0.4}$)

ϵ	Inference procedure			
	$BFDR_{0.25}$ (%)	$BFDR_{0.40}$ (%)	SimBaS (%)	PWCI (%)
0.001	0.00	0.00	1.20	14.7
0.005	0.00	0.00	1.90	19.3
0.015	0.00	0.00	4.03	25.1
0.025	0.00	0.00	4.24	28.6

All inference procedures use $\alpha = 0.05$

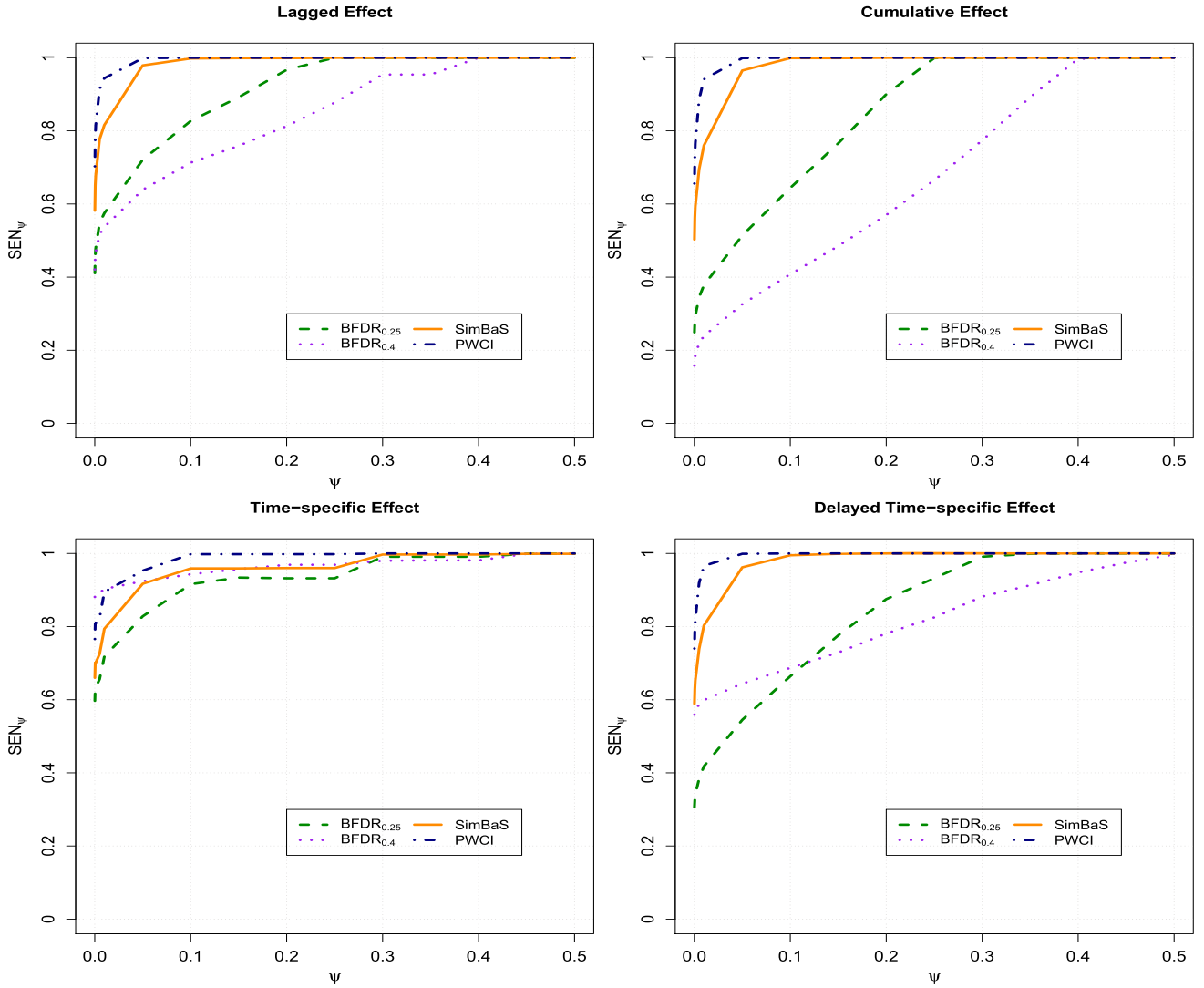

Fig. 3 Sensitivity, SEN_ψ, for the SimBaS, point-wise credible intervals (PWCIs), and BFDRs at $\delta = 0.25$ ($BFDR_{0.25}$) and 0.4 ($BFDR_{0.4}$) applied to the general relationships. This figure appears in color in the electronic version of this article

Under the peak setting, the sensitivity of the 95% PWCI is the fastest to one with the SimBaS procedure in second. Figure 3 depicts a similar story for the other surfaces when $N = 50$ and $T = V = 64$. SEN_ψ of the BFDR is highly dependent upon surface type, size of δ, sample size, and sampling density. FDR_ϵ in the peak setting is smallest for the BFDRs, regardless of δ, across all ϵ with SimBaS having the second smallest value. The PWCIs have high FDR for even modest ϵ which is also the case in the general relationships. The values of FDR_ϵ are sensitive to sample size and sampling density for the BFDRs, reaching as high 40% in some settings for the smallest ϵ. Tables of FDR_ϵ values for settings not presented here are in Supplementary Material. In contrast, SimBaS has low FDR_ϵ with SEN_ψ that is comparable to that of PWCI.

5 Analysis of journeyman boilermaker data

Harezlak et al. (2007) and Cavallari et al. (2008) analyze data from 14 journeyman boilermakers relating measured standard deviation of the normal-to-normal intervals, or SDNN, to each journeyman's microenvironment pollution exposure over the course of the day. Pollution exposure was taken from personal respirators that recorded the level of exposure to particulate matter finer than $2.5\mu g$ in diameter. Both SDNN and $PM_{2.5}$ were measured over the course of the workday and aggregated to the five-minute scale. Exposure to $PM_{2.5}$ primarily came from two sources: residual oil fly ash or ROFA, a by-product of the manufacturing process, and cigarette smoke during mandatory breaks. Particulate matter as fine as $PM_{2.5}$ is known to be detrimental to heart health (see for example the MESA study, Kaufman et al. (2016) and references therein); thus, it is important to quantify the effect of occupational exposure to $PM_{2.5}$ on heart health. The study of SDNN helps provide insights into the biological mechanisms underlying these observed health effects. Harezlak et al. (2007) find elevated levels of association during the morning hours that corresponded to mandatory breaks, which were then followed by depressions in the estimated surface later in the day. To try to better understand these spikes in association, we focus our analysis on the first three hours of the workday beginning at 8:30 am and going until 11:30 am. Unlike Harezlak et al. (2007) whose method does not yield inference on the estimated surface, we estimate joint intervals and calculate both SimBaS and the BFDR to determine the significance of these peaks and troughs.

Prior to analysis, we log-transform and center and scale both SDNN and $PM_{2.5}$. Thus, changes in the estimated surface correspond to one standard deviation changes in the log of $PM_{2.5}$ and result in changes in the log of SDNN. The sampled curves are spiky and irregular; see Supplementary Material for graphical depictions. To investigate the morning hours only, we take $T = V = 34$ measurements. As in the simulation, we use $J = 3$ levels of decomposition for both wavelet-packet transformations and select Daubechies wavelets with 3 vanishing moments. Given the results of the simulation, we retain only the first two scales of wavelet-packet coefficients in $\mathbf{X}^{W_P}$, resulting in only 25% of the columns being kept. We allow this model to run longer than the simulated settings, taking 1000 posterior samples after a burn-in of 3000 and monitor convergence using the potential scale reduction factor (Gelman and Rubin 1992) and trace plots; see Supplementary Material for both.

Figure 4 contains the posterior surface estimate (top left panel) for the journeyman data alongside the lower (top right panel) and upper (bottom panel) bounds of the joint credible intervals. Given the centering, the estimated intercept is nearly zero for all t. We present this estimated intercept and its associated credible intervals in Supplementary Material. In Fig. 4, moving across the v-axis from left to right, we see a pattern of time-specific depressions at the start of the workday followed by elevations, which culminate in a delayed time-specific effect around $v = 8$ and $t = 20$. This suggests that initial exposure to ROFA is associated with a delayed decrease in SDNN. Further, another exposure later in the day is associated with elevated SDNN that is sustained until, and peaks at, $t = 20$. After this peak, there is another depression that occurs around $v = 16$ and $t = 20$ suggesting a decrease in SDNN associated with an exposure approximately 2–4 measurements prior. These patterns are consistent with the analysis in Harezlak et al. (2007), but using our approach we are able to show they are also significantly different from zero as suggested by the upper and lower bounds.

Given its performance in simulation, we turn to the SimBaS procedure to confirm these inferential observations. Figure 5 displays the SimBa scores (left panel) and whether or not the score is less than $\alpha = 0.05$ (right panel). By design, SimBa scores are capped at 0.5 as evaluation of higher α values is not meaningful: Regions equal to 0.5 suggest the corresponding coefficients require a joint credible interval with $\alpha > 0.5$ to exclude zero. We see that the smallest SimBa scores correspond to the regions of both increased and decreased association in the estimated surface from Fig. 4. The right panel of Fig. 5 indicates that these regions are also significantly different from zero. The BFDR, evaluated at $\delta = 0.25$ and 0.4 and graphically presented in Supplementary Material, further confirms the significance of these regions.

To assess the choices to retain only 25% of the coefficients and to use $J = 3$ levels of decomposition, we conduct a sensitivity analysis that varies both. In the above model, which retains only 25% of the $\mathbf{X}^{W_P}$ coefficients, $R^2_{W,\text{ave}}$ is 34.16%. The model that retains 50% of the coefficients only modestly increases $R^2_{W,\text{ave}}$ to 37.51%. The choice of levels of decomposition represents a trade-off in the smoothness of the estimate and the number of coefficients available within each scale. To assess the impact of choice of J, we also fit the model with $J = 4$ levels of decomposition. In the wavelet-space, the model that retains 25% of the coefficients results in a $R^2_{W,\text{ave}}$ of 21.75%. Finally, the $J = 4$ model that retains 50% of the coefficients has an $R^2_{W,\text{ave}}$ of 17.13%. While the model we present does not have the largest $R^2_{W,\text{ave}}$, the trade-off of $\sim$3% gain in $R^2_{W,\text{ave}}$ versus the additional computation time and performance in the peak setting suggests that the model that retains only 25% of coefficients is preferred. Increasing the levels of decomposition only decreases $R^2_{W,\text{ave}}$.

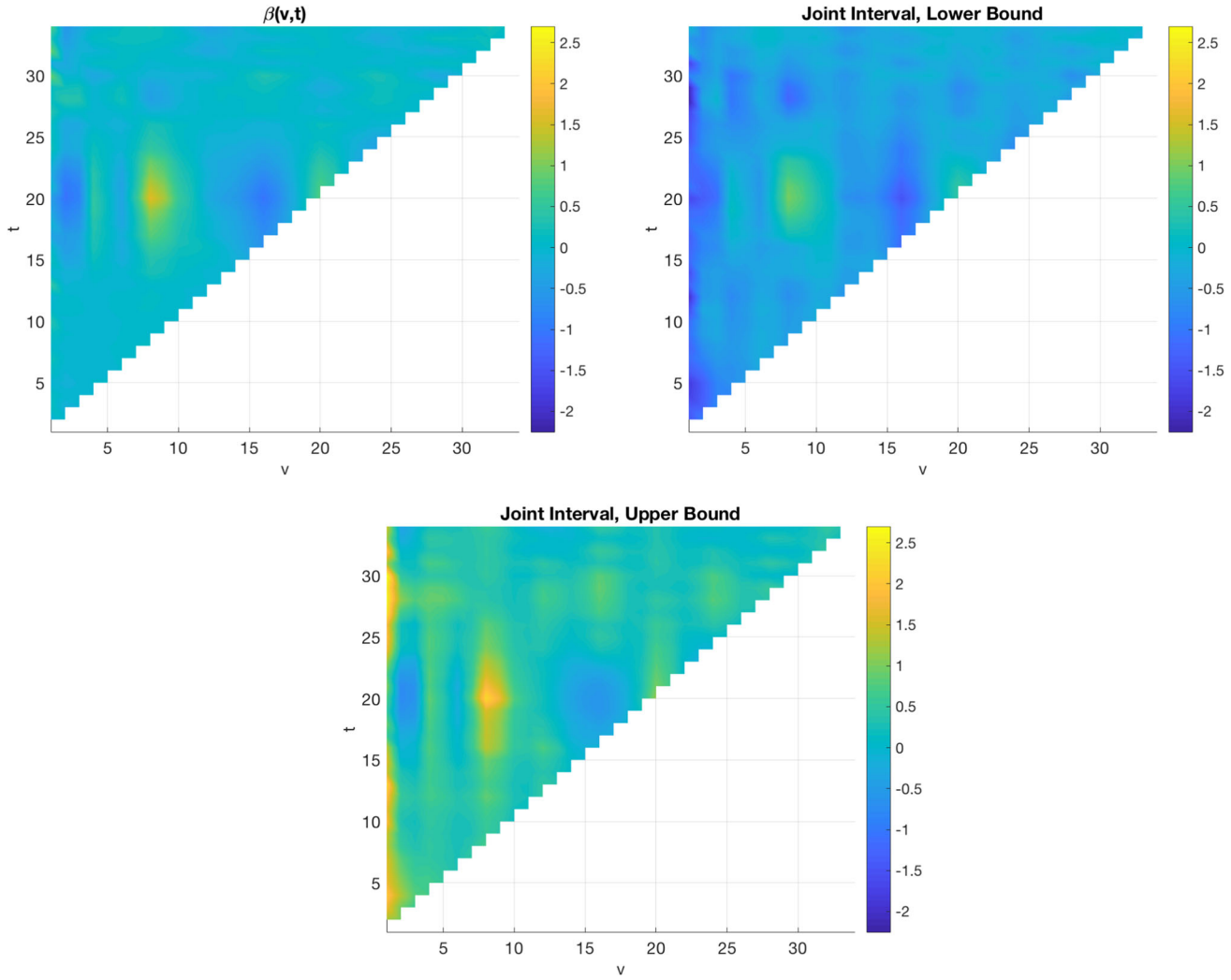

Fig. 4 Estimated surface of association (upper left panel) between log(SDNN) and log($PM_{2.5}$) in journeyman boilermakers during the morning, 8:30 am to 11:30 am. Lower (upper right panel) and upper (lower panel) bounds of the joint interval. This figure appears in color in the electronic version of this article

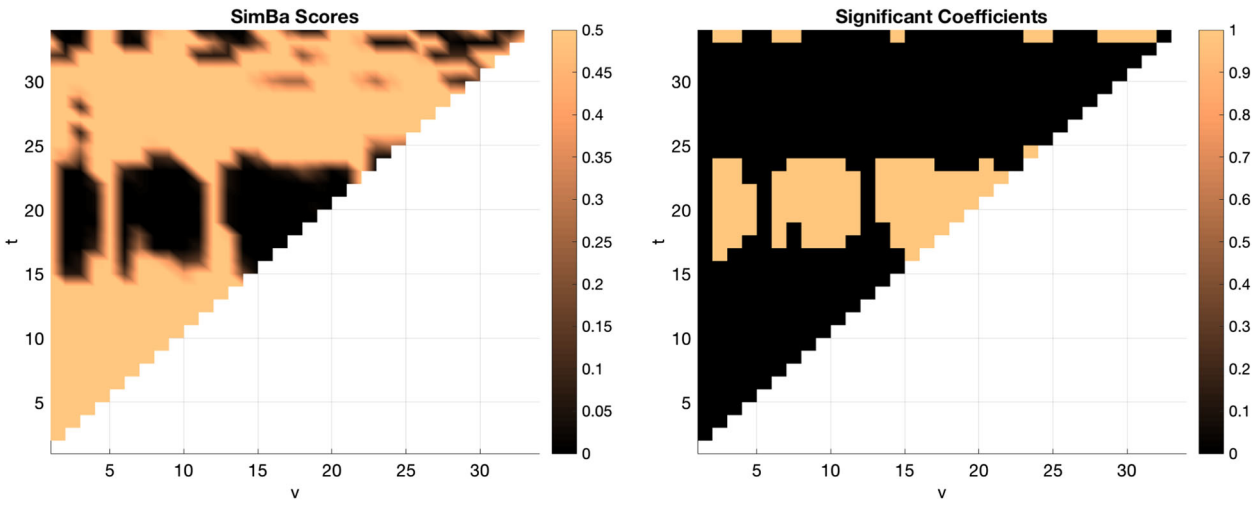

Fig. 5 SimBa scores (left panel) and corresponding significant coefficients (right panel) for the surface of association. Significance is determined using an $\alpha = 0.05$ cutoff. This figure appears in color in the electronic version of this article

6 Discussion

One of the difficulties of estimating a historical functional effect is maintaining the constraint which previous authors have used tent-shaped basis functions to achieve. The existing wavelet-based methodology cannot estimate historical effects. However, wavelet-packet coefficients from a DWPT have an exploitable relationship between their location indices and the original time scale that allows us to sample only the desired coefficients in the wavelet-packet space via our prior specification. The constraint is then maintained when projecting back to the data space. In this work, we show that when performing the DWPT on both $y(t)$ and $x(v)$, we can use wavelet-packet basis functions and a historical constraint prior to model historical effects in the wavelet domain.

The current literature on HFLMs is limited, with a focus on the estimation of surface effects and the determination of model fit criterion. Most methods implement spline-based basis functions. While one method does allow for different basis expansions, the authors only present results for spline-based models. Further, inference procedures are not discussed in the existing literature. To our knowledge, the model we present here represents the first work in wavelet-based modeling of historical effects as well as the first Bayesian HFLM. Additionally, our method employs a novel use of wavelet-packets, which have not previously been used in functional regression models. We also adapt several established multiplicity adjusted Bayesian inferential procedures to the constrained surface to generate posterior intervals and identify peaks and regions of significant coefficients. Finally, we conduct a formal comparison of our approach to two existing methods.

We demonstrate that, in simulation, the wavelet-packet HFLM can accurately estimate several realistic historical surface settings. In particular, we show that, regardless of the percent of $\mathbf{X}^{W_P}$ coefficients we retain, the wavelet-packet HLFM has good RMISE levels that are similar across sample size and sampling density. Our approach outperforms both the FDBoost and FEB-based methods in terms of estimation for the more general relationships, regardless of the number of coefficients retained. Under the data-motivated setting, the model retaining 25% of the wavelet-packets performs best. Further, we show that the joint credible intervals provide better coverage than PWCIs. For identifying significant coefficients, the SimBaS procedure strikes a balance between lower FDR_ϵ while maintaining reasonably high sensitivity without having to select a δ value. Given the sensitivity to the choice of δ in the BFDR and the lowered coverage of the PWCI, we propose the use of joint intervals and the SimBaS procedure for inference. We also recommend retaining only 25% of the wavelet-packet coefficients given its performance in simulation and reduced computational burden. Finally, we apply the proposed model to analyze data on the association between HRV and $\text{PM}_{2.5}$ exposure in a panel of journeymen boilermakers, focusing on the first three hours of exposure during the workday. Using the wavelet-packet HFLM, we are able to not only estimate regions of association but clearly identify them as representing significant changes in SDNN using SimBaS. The reasonable $R^2_{W,\text{ave}}$ for the model retaining only 25% of the coefficients further supports our recommendation.

References

Brockhaus, S., Melcher, M., Leisch, F., Greven, S.: Boosting flexible functional regression models with a high number of functional historical effects. Stat. Comput. **27**, 913–926 (2017)

Cavallari, J.M., Fang, S.C., Eisen, E.A., Schwartz, J., Hauser, R., Herrick, R.F.: Time course of heart rate variability decline following particulate matter exposures in an occupational cohort. Inhal. Toxicol. **20**, 415–422 (2008)

Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**, 457–511 (1992)

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J., Harezlak, J., McLean, M.W., Swihart, B., Xiao, L., Crainiceanu, C., Reiss, P.T.: refund: Regression with Functional Data. R package version 0.1-21 (2019). https://CRAN.R-project.org/package=refund

Harezlak, J., Coull, B.A., Laird, N.M., Magari, S.R., Christiani, D.C.: Penalized solutions to functional regression problems. Comput. Stat. Data Anal. **51**, 4911–4925 (2007)

Ivanescu, A.E., Staicu, A.-M., Scheipl, F., Greven, S.: Penalized function-on-function regression. Comput. Stat. **30**, 539–568 (2015)

Kaufman, J.D., Spalt, E.W., Curl, C.L., Hajat, A., Jones, M.R., Kim, S.-Y., Vedal, S., Szpiro, A.A., Gassett, A., Sheppard, L., Daviglus, M.L., Adar, S.D.: Advances in understanding air pollution and cardiovascular diseases: the multi-ethnic study of atherosclerosis and air pollution (MESA Air). Global Heart **11**, 343–352 (2016)

Kim, J.S., Staicu, A.-M., Maity, A., Carroll, R.J., Ruppert, D.: Additive function-on-function regression. J. Comput. Graph. Stat. **27**, 234–244 (2018)

Kim, K., Şentürk, D., Li, R.: Recent history functional linear models for sparse longitudinal data. J. Stat. Plan. Inference **141**, 1554–1566 (2011)

Magari, S.R., Hauser, R., Schwartz, J., Williams, P.L., Smith, T.J., Christiani, D.C.: Association of heart rate variability with occupational and environmental exposure to particulate air pollution. Circulation **104**, 986–991 (2001)

Malfait, N., Ramsay, J.O.: The historical functional linear model. Can. J. Stat. **31**, 115–128 (2003)

Malloy, E.J., Morris, J.S., Adar, S.D., Suh, H., Gold, D.R., Coull, B.A.: Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. Biostatistics **11**, 432–452 (2010)

Meyer, M.J., Coull, B.A., Versace, F., Cinciripini, P., Morris, J.S.: Bayesian function-on-function regression for multi-level functional data. Biometrics **71**, 563–574 (2015)

Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.-M.: Wavelets and their applications. ISTE Ltd, London (2007)

Morris, J.S.: Functional regression. Ann. Rev. Stat. Appl. **2**, 321–359 (2015)

Morris, J.S., Brown, P.J., Herrick, R.C., Baggerly, K.A., Coombes, K.R.: Bayesian analysis of mass spectrometry proteomic data using

wavelet-based functional mixed models. Biometrics **64**, 479–489 (2008)
Morris, J.S., Carroll, R.J.: Wavelet-based functional mixed models. J. R. Stat. Soc. Ser. B **68**, 179–199 (2006)
Nason, G.P.: Wavelet methods in statistics with R. Springer, Berlin (2008)
Percival, D.B., Walden, A.T.: Wavelet methods for time series analysis. Cambridge University Press, Cambridge (2000)
Pomann, G.-M., Staicu, A.-M., Lobaton, E.J., Mejia, A.F., Dewey, B.E., Reich, D.S., Sweeney, E.M., Shinohara, R.T.: A lag functional linear model for prediction of magnetization transfer ratio in multiple sclerosis lesions. Ann. Appl. Stat. **10**, 2325–2348 (2016)
Ruppert, D., Wand, M.P., Carroll, R.J.: Semiparametric regression. Cambridge University Press, Cambridge (2003)
Scheipl, F., Greven, S.: Identifiability in penalized function-on-function regression models. Electron. J. Stat. **10**, 495–526 (2016)
Scheipl, F., Staicu, A.-M., Greven, S.: Functional additive mixed models. J. Comput. Graph. Stat. **24**, 477–501 (2015)
Zhu, H., Brown, P.J., Morris, J.S.: Robust, adaptive functional regression in functional mixed model framework. J. Am. Stat. Assoc. **106**, 1167–1179 (2011)